BBC RD 1976/11

BBC

RESEARCH DEPARTMENT

1969 1974
THE QUEEN'S AWARD TO INDUSTRY

REPORT

An experimental comparison of four methods for 64 kbit/s coding of speech with a 7 kHz bandwidth

D.W. Stebbings, B.A.(Cantab.)

Research Department, Engineering Division
THE BRITISH BROADCASTING CORPORATION May 1976

BBC RD 1976/11
UDC 534.86:
681.846

AN EXPERIMENTAL COMPARISON OF FOUR METHODS FOR 64 KBIT/S CODING OF SPEECH WITH A 7 KHZ BANDWIDTH

D.W. Stebbings, B.A. (Cantab.)

Summary

Four possible methods are investigated for coding speech signals with a 7 kHz bandwidth to give commentary-circuit programme-quality, within a digital telephony bit-rate of 64 kbit/s. They all exploit redundancy in speech signals, and produce different subjective effects. Apparatus was constructed, either implementing or closely simulating each method, and tests were made with a wide variety of speech and music in order to assess the relative merits of each method. It was concluded that two of the methods gave marked improvements over techniques previously studied. The two preferred methods respectively used pitch-halving and variable sampling-rate bit-rate reduction techniques; they both gave speech quality which may be acceptable for commentary circuits. None of the methods would be acceptable for musical items.

Issued under the authority of

P. Rainger

Head of Research Department

Research Department, Engineering Division,
BRITISH BROADCASTING CORPORATION

May 1976

(EL-113)

AN EXPERIMENTAL COMPARISON OF FOUR METHODS FOR 64 KBIT/S CODING OF SPEECH WITH A 7 KHZ BANDWIDTH

AN EXPERIMENTAL COMPARISON OF FOUR METHODS FOR 64 KBIT/S CODING OF SPEECH WITH A 7 KHZ BANDWIDTH

D.W. Stebbings, B.A. (Cantab.)

1. Introduction

A large number of BBC programme contributions are currently made by reporters, correspondents and commentators, both at home and abroad, using the public telephone network. The sound-programme quality of these contributions is normally well below that obtained over circuits dedicated to sound programme, e.g. 'music lines'. With the planned use by the UK Post Office of digital telephone circuits having a digital rate of 64 kbit/s, the possibility exists of the BBC using such circuits for programme contributions. By the mid 1980's it is likely that such circuits may become widespread throughout Europe and the developed countries in the world, and access by broadcasters to the circuits should be practicable. Unfortunately, the quality attainable with a standard telephone installation even on 64 kbit/s circuits is low by broadcasting standards, partly because of the limited bandwidth determined by the handset and 8 kHz sampling frequency, and partly because the coding method uses only 8 bits per sample and A-law companding,[1] so that non-linear distortion contributes significantly to the relatively poor quality.

The purpose of the work described in this Report was to seek alternative methods of coding which would widen the passband to the limits 30 Hz and 7·4 kHz approximately, with the same bit-rate, and which would for speech transmission give negligible or small audible defects in the received signal. To double the transmitted bandwidth whilst maintaining a given signal-to-noise ratio implied doubling the bit-rate, and therefore some redundancy in the speech signals had to be exploited in a way that would allow the bit-rate to be kept at 64 kbit/s.

Although there has been considerable effort by many research establishments on reducing the bit-rate necessary to transmit speech, the requirements for various communication systems differ greatly. For instance, military requirements are usually for a speech system with very low bit-rates, of the order of 10 kbit/s, able to withstand very high bit-error rates, say up to 1 in 10^2, where the quality of the received signal is unimportant, provided that intelligibility is good. This system contrasts with a possible system for broadcasting, where the requirement is for a system that may use up to 64 kbit/s to give good signal quality when transmitted over links having relatively low error rates. Therefore, it is not surprising that little work had been done which was directly applicable. In 1973 work was begun on low bit-rate speech transmission in BBC Research Department; a relevant consequence of the first phase of this work[2] is discussed in Section 2.

This Report describes four experimental methods for 64 kbit/s coding of speech with a 7 kHz bandwidth; the results for each method were tape-recorded using various types of programme material. Extracts were taken from recordings of Harvard speech sentences,[3] an electronic gong and mixed items of music. As a starting reference (lower bound) in quality, recordings were made using an 8-bit A-law digital telephony system, and as an ideal goal (upper bound) in quality, the original high-quality programme signal was reproduced band-limited to 7·4 kHz.

2. Separate description of envelope and zero crossings of the top octave-band

Previous BBC research[2] inspired a proposal* that further study be made of a method in which the envelope and zero-crossings of the top octave-band of a speech signal are isolated and separately coded for transmission. This method offered a promising way of obtaining the required bit-rate. In the earlier work it was found, in a simulation, that the rate of change of envelope in the upper octave, here 3·4 kHz to 6·8 kHz, was such that the envelope could be transmitted in a bandwidth of approximately 750 Hz without detriment to sound quality. Because of this allowable bandwidth restriction, the bit-rate necessary to transmit the envelope was about one quarter of that apparently required. However, in the simulation neither the envelope signal nor the zero-crossing signal was digitally encoded. It was estimated that only 24 kbit/s would be needed for the upper octave if it were processed by this method, which would leave an apparently adequate 40 kbit/s, say five bits per sample with near-instantaneous (n.i.) companding[4] and 8 kHz sampling frequency, for the signal in the band up to 3·4 kHz. At that time no clear method of encoding the top-octave zero-crossings had been worked out. Since then it has become clear that the zero-crossing signal cannot be sufficiently band-restricted to afford accommodation of both the coded envelope and zero-crossing signals within the estimated 24 kbit/s available. Thus, this approach did not fulfil its earlier promise, and further work on it ceased.

3. Sub-Nyquist sampling

As is well known, to avoid unwanted alias components of a sampled signal overlapping the wanted baseband signal, the sampling rate should be at least twice that of the highest frequency in the baseband signal. However, if the required band of frequencies (or an upper part of this band) is initially comb-filtered, then it is possible to choose a particular sampling frequency such that the alias components fall in the gaps of the spectrum produced by the comb-filter. In a decoder the alias components may be removed by a second comb-filter similar to the first. The problem of using this technique is that comb-filters have a considerable deleterious effect on audio quality. It was found, however, that if the action of the filter was confined to

* The proposal was made by M.G. Croll.

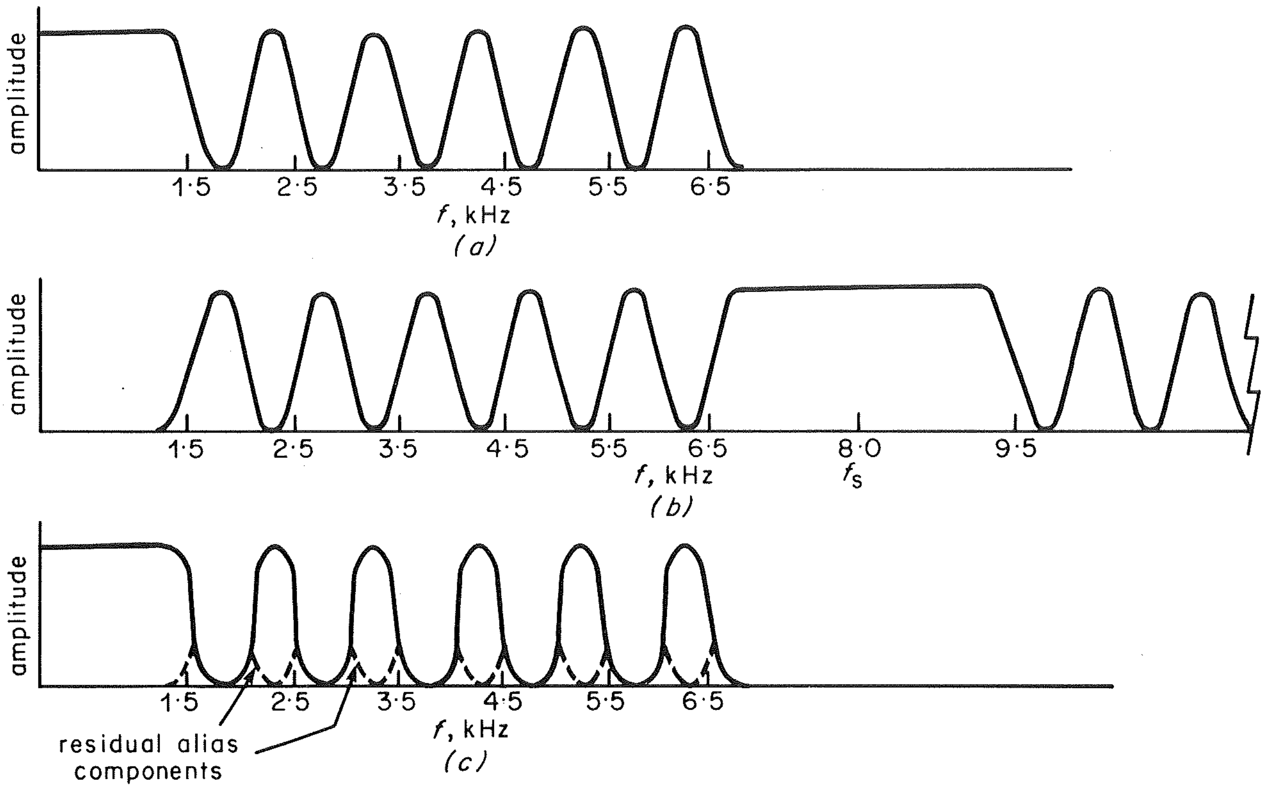

Fig. 1 - Spectra produced by a comb-filter and sub-Nyquist sampling

(*a*) Baseband after first comb-filter (second order)

(*b*) Alias components from sampling (*c*) Baseband after second comb-filter, showing residual alias components

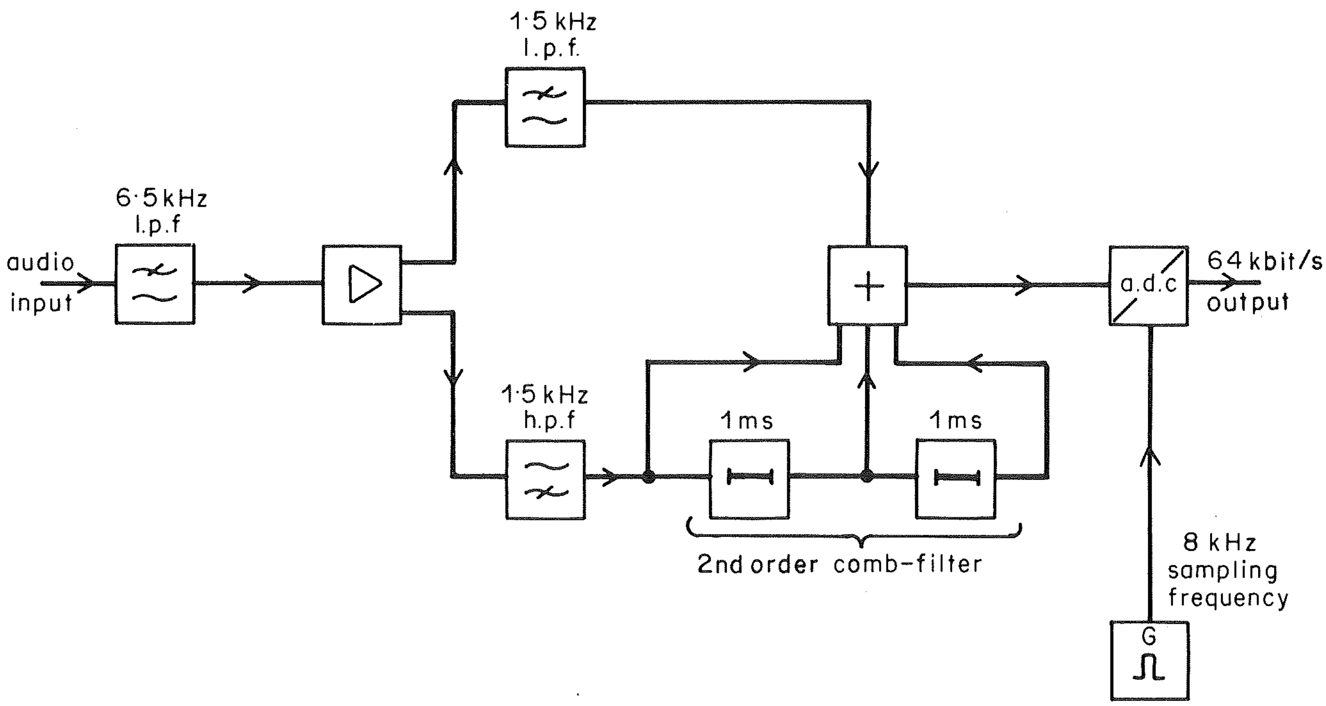

Fig. 2 - Block diagram of a possible arrangement for a 'sub-Nyquist' coder

the upper two octaves of the speech signal, where potential bit-saving is high, there is then much less effect on programme quality. Fig. 1 illustrates spectra produced by this method. Fig. 1(*a*) shows the baseband signal after the first comb-filter and Fig. 1(*b*) shows the alias components generated from sampling. With practical second order comb-filters, small groups of alias components of maximum amplitude 12 dB below peak-programme level remain in the recovered signal. Fig. 1(*c*) illustrates the baseband signal after the second comb-filter and shows these residual alias components. If higher-order comb-filters are used the residual alias components are reduced in amplitude.

Fig. 2 gives a block diagram of an experimental digital coder that could be used for this method. A similar band-split and comb-filter is used in the decoder together with a 7 kHz low-pass filter. In the laboratory this technique was simulated.* The saving in bit-rate is about 50% assuming a sampling rate of 8 kHz instead of 16 kHz for a recovered audio bandwidth of 6·5 kHz. After several informal listening tests seeking a minimum effect on audio quality, it was found that the spacing between the 'teeth' of the comb-filter response was best at about 1 kHz; values between 160 Hz and 2 kHz were tried.

4. Bit-rate reduction by pitch-halving

Many research workers have studied methods to vary the syllabic rate of speech and yet maintain the pitch.[5] These methods all rely on being able to remove and repeat blocks of samples of the speech waveform. It can be seen by simple inspection of waveforms that adjacent speech waveform segments, of about 10 to 30 ms duration corresponding to pitch, are very often similar to each other; transmission of alternate segments only (effectively halving the pitch) is then sufficient, provided that the transmitted segments (after restoration of the pitch) are each repeated in the receiver. In practice, of course, significant differences between blocks of samples do occur, and distortion of the output signal then arises when blocks are omitted or repeated.

Exploiting this redundancy in speech, a digital coder, which removed alternate 16 ms segments of the signal, and a decoder, which repeated each received segment, were constructed as outlined in Fig. 3; near-instantaneous companding[4] was used, by means of a simulator,[6] to reduce the bit-rate further. The problem with segment repetition was that the audio output became modulated at the repetition rate, causing a low-frequency buzz, with higher frequency clicks when there were large discontinuities between adjacent samples. It was found however, that if the action of this system was confined to the signal above 1·7 kHz, then the speech distortion was considerably reduced. Furthermore, by using pre- and de-emphasis of the higher frequencies, before and after the coder and decoder, respectively, the audibility of the high-frequency distortion was further reduced.

* G.R. Mitchell constructed the apparatus necessary for this simulation and the other experiments described in this Report.

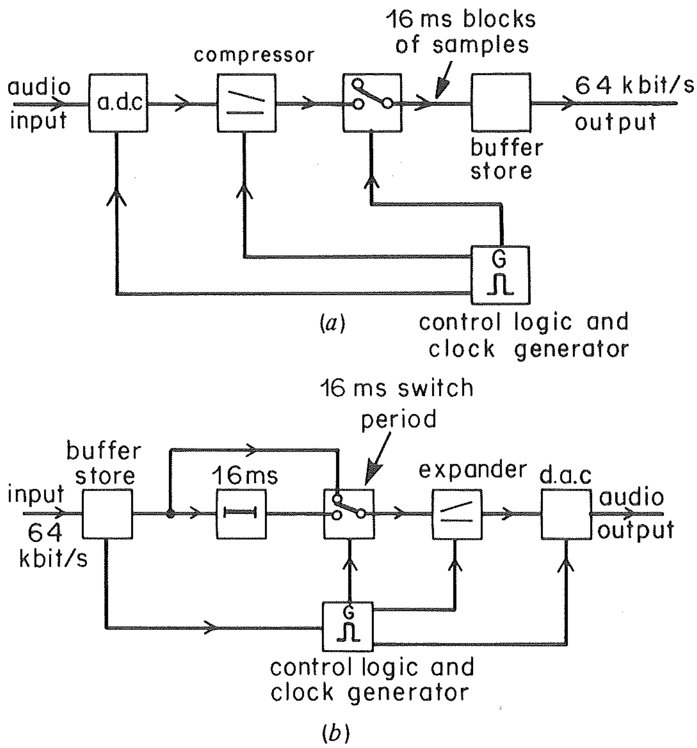

Fig. 3

(*a*) Coder which removes alternate 16 ms blocks of samples
(*b*) Decoder which reproduces 16 ms blocks of samples twice

Finally, since the method omits and repeats segments of fixed duration, say n ms, then at frequencies which are near integral multiples of $1/n$ little distortion occurs; conversely, distortion arising from frequencies which are near odd multiples of $1/2n$ is great, as shown diagrammatically in Fig. 4. A comb-filter was therefore incorporated in the coder which removed signal components at the frequencies suffering gross distortion. Fig. 5 shows the final arrangement of a possible coder with the 1·7 kHz high-pass filter, the high-frequency pre-emphasis and the comb-filter. A tape-recording was made of one frequency band at a time, which simulated bit-rates of 6 bits per sample, n.i. companded, for the signal up to 1·7 kHz, corresponding to a bit-rate of 24 kbit/s; and 5 bits per sample, n.i. companded, corresponding to a bit-rate of 40 kbit/s for the signal above 1·7 kHz, giving a total bit-rate of 64 kbit/s.

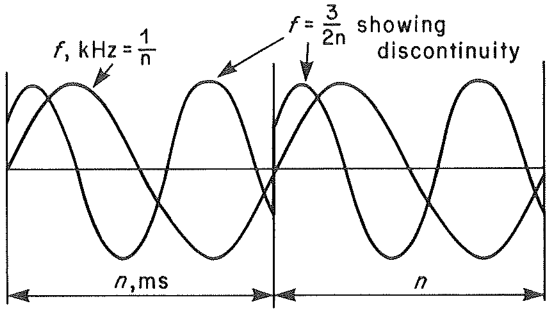

Fig. 4 - Regular duplication of blocks of samples causing severe distortion in certain groups of frequencies, and little or no distortion in other groups

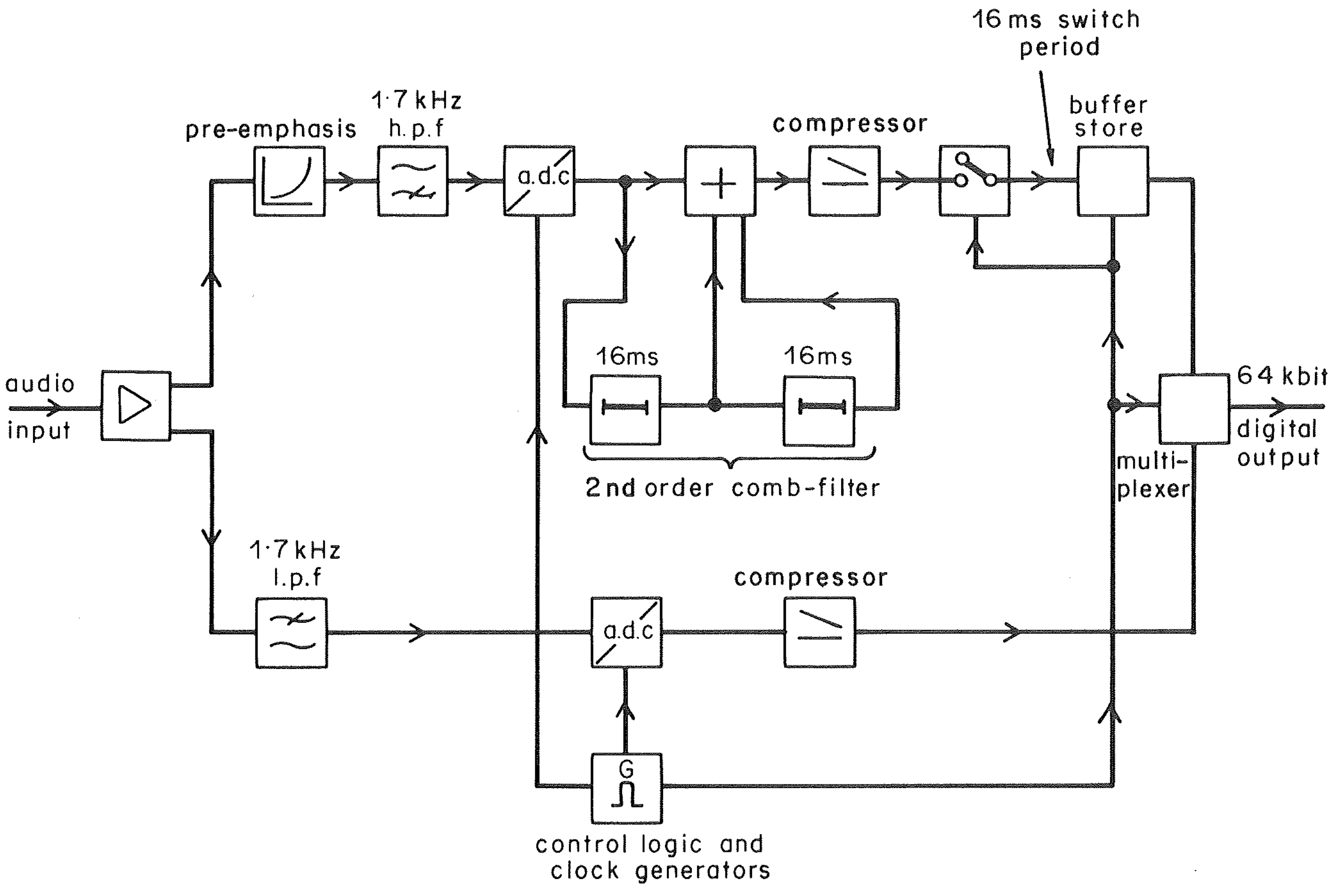

Fig. 5 - Block diagram of a coder for the pitch-halving method

As before, the compandor simulator[6] was used to make the tape recordings of the effect of the system on programme.

The results of informal listening tests with speech signals using this method were encouraging, although distortion was audible with some voices, and the effect of the periodicity of the comb-filter was also audible.

Many listeners felt that this 'pitch-halving' method was the best one of the four investigated. Subjective results with music were not good; piano and violin were grossly distorted, and the system would be totally unsuitable for music transmission.

5. Variable sampling-rate using a constant bit-rate

It is generally necessary to sample a baseband signal at a rate at least twice that of the highest frequency present in that signal. For example, if a word is spoken which contains no components above 2 kHz in frequency, then sampling at a little over 4 kHz would be sufficient to fully describe that signal. If the next word spoken has components up to 4 kHz, then sampling at a little over 8 kHz would be sufficient. If the sampling rate were to be varied in accordance with the speech-spectrum variations, and if a constant quantising accuracy were used, then the overall bit-rate would be proportional to the sampling rate. Although we have here a basis for bit-rate reduction, difficulties could occur when transmitting the data over digital paths, most of which are normally designed to accept constant bit-rates. To overcome this problem, whilst conserving the benefit of a variable sampling-rate, a method has been examined in which the quantising accuracy is reduced as the sampling rate is increased, in such a way that a constant overall bit-rate is produced. In a practical system, the sampling rate would be fixed for blocks of samples, lasting a few milliseconds, i.e. a fraction of one syllable of speech.

A simulation of the method was made in which the variable sampling rate could have one of four values. These were 16 kHz, 10·5 kHz, 8 kHz and 6·2 kHz. The corresponding bits per sample would be 4, 6, 8 and 10, to give an approximately constant bit-rate of 64 kbit/s. To determine the appropriate choice of sampling frequency, four low-pass filters and three signal-level threshold detectors were used.

Fig. 6 is a block diagram of the simulated coder, incorporating the threshold detectors, low-pass filters and control logic. Frequency components up to 2·9 kHz (corresponding to 10 bits/sample and therefore high quantising accuracy) were added directly to the output signal. When signals higher than 2·9 kHz in frequency appeared at the inputs of the three threshold detectors, the logic switched to the sampling rate corresponding to the low-pass filter of the highest frequency for which the level-threshold had been exceeded. The compandor simulator[6] was again

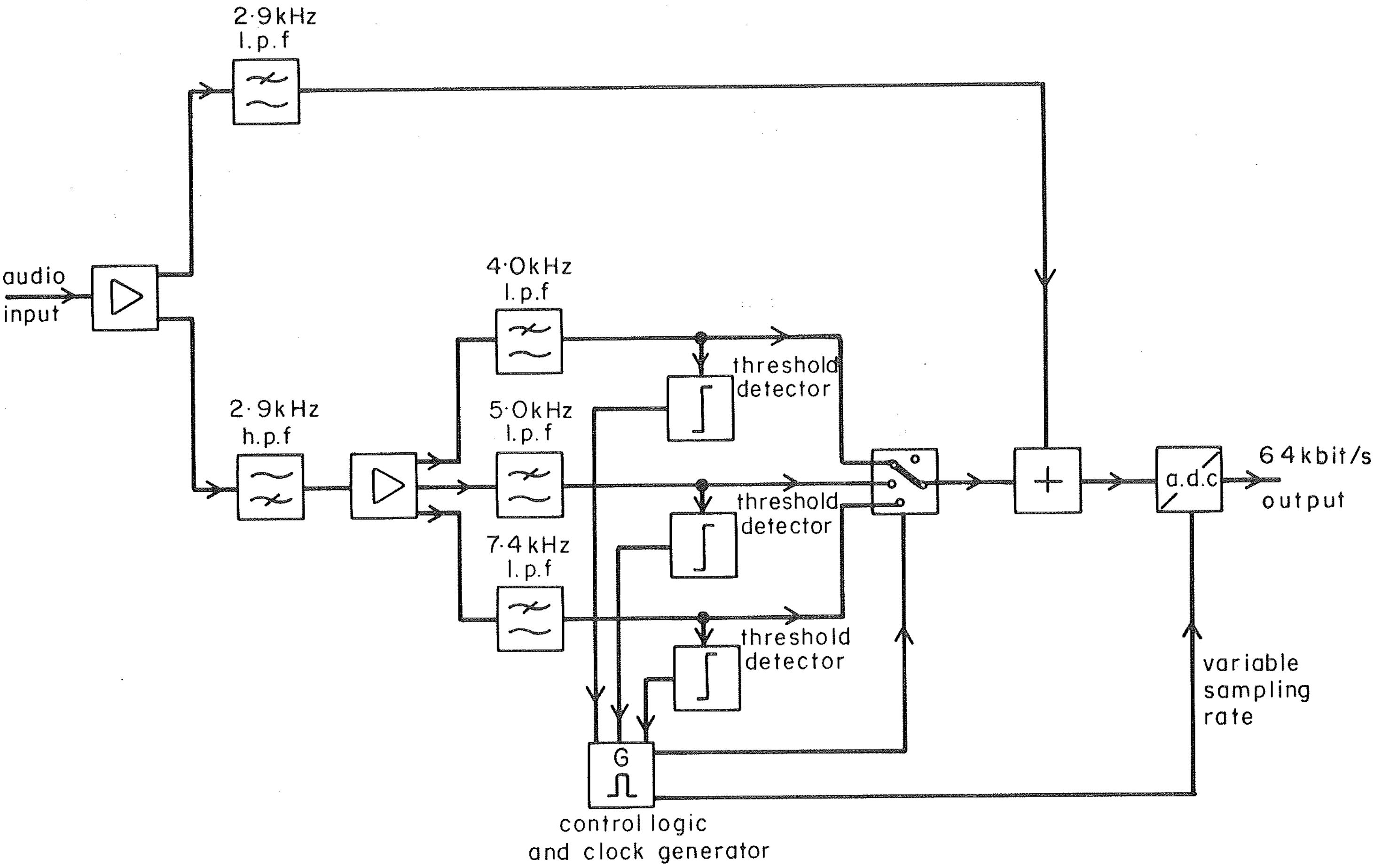

Fig. 6 - Block diagram of a variable sampling-rate coder

used to simulate the variation in quantising accuracy, corresponding to the variable sampling-rate with a constant bit-rate, and tape recordings were made of the results of this method on the same speech and musical excerpts used with the other three methods.

It was felt initially that low-level, high-frequency signals might be important for overall clarity of the reproduced voice, even though the mean energy at these frequencies would be low. With high-level, high-frequency sounds, the reduction in quantising accuracy and the corresponding increase in quantising noise would tend to be masked by the high-frequency content of the signal. However, statistical studies of the spectra of speech signals showed that most of the speech energy is concentrated at the lower end of the frequency spectrum. In this system therefore, low sampling-rates (and therefore high quantising-accuracy) were in operation for most of the time.

On listening tests the clarity of the speech was good, and earlier fears that the loss of low-level high-frequency sound would cause objectionable effects were largely unfounded. However, the main defect with this method was that it was possible to hear clearly the effect of the varying bandwidth as the sampling rate was varied. This produced, for instance, variation in the spectrum and audibility of the background noise, and some observers rated these effects more disturbing than the repetitive distortion heard with the pitch-halving method outlined in Section 4.

6. Conclusions

Of four speech-coding methods examined, two, the pitch-halving and the variable sampling-rate methods, gave better quality than previously known methods of coding 7 kHz bandwidth speech to give a 64 kbit/s signal. There was a marked improvement in the clarity of the received speech compared with normal telephone quality. Unfortunately, both methods gave side effects which were subjectively obvious; the pitch-halving gave periodic distortion, and the variable sampling-rate method gave audible bandwidth-change effects. Nevertheless they both gave speech quality which may be acceptable for commentary circuits. The pitch-halving method would be considerably simpler to engineer as a final system, and it would not require the additional complication of the bits required for signalling the sampling rate, which the variable sampling-rate method requires.

A further possible application of 64 kbit/s transmission systems is for music signals up to a bandwidth of 5·5 or 6 kHz, as a possible feed for m.f. and h.f. broadcasting

transmitters. Unfortunately, the quality of the music processed by both of the preferred methods examined here was poor, and would be inadequate even for such low-bandwidth music-signal distribution.

It is recommended that no further studies be made of any of the four methods described in this Report; a better method should be sought.

7. References

1. RICHMAN, G.D. 1971. Construction of p.c.m. laboratory equipment for speech transmission studies. UK Post Office Research Department Report No. 229.

2. CROLL, M.G. 1975. The possibility of sending programme speech contributions digitally over the public telephone network. BBC Research Department Report No. 1975/6.

3. IEEE, 1969. 1965 revised list of phonetically balanced sentences (Harvard Sentences); IEEE Recommended Practice for speech quality measurements. *IEEE transactions on Audio and Electroacoustics*, 1969, September, pp. 239 – 246.

4. CROLL, M.G., MOFFAT, M.E.B. *and* OSBORNE, D.W. 1973. 'Nearly instantaneous' digital compandor for transmitting six sound-programme signals in a 2·048 Mbit/s multiplex. *Electronics Letters*, 1973, **9**, 14, pp. 298 – 300.

5. SPRINGER, A.M. 1957. 'Acoustical time regulator'. *Journal of the Acoustical Society of America*, March 1957, **29**, pp. 341 – 343.

6. OSBORNE, D.W. 1972. Digital sound signals: further investigation of instantaneous and other rapid companding systems. BBC Research Department Report No. 1972/31.

SMW/VY

Printed by BBC RESEARCH DEPARTMENT, Kingswood Warren, Tadworth, Surrey, KT20 6NP